The Problem-Oriented Medical Synopsis: Coding, Indexing, and Classification Sub-Model

Frank W. Stitt MD, Departments of Epidemiology and Medicine
University of Miami School of Medicine, Miami, Florida

A clinical information system consists of four major components: the clinical database, decision support, data analysis (including outcomes), and the development system. We have created such a system using generally available database methodology. The system is documented using a conceptual model, a physical model, and sub-models for individual components. A key sub-model of the the clinical database, for record-keeping, has been defined for coding, indexing, and classification of the medical narrative typically encountered in medical records. We describe an approach to the development of the coding component that results in a hybrid system for recording information, locating indexed information, and summarizing it for analysis of outcomes. These are based on a primary term list—the problem glossary; SNOMed—the Systematized Nomenclature of Medicine (3rd. edition); and ICD-9-CM. The relationship with the UMLS is also discussed.

INTRODUCTION

We have developed a clinical database called the Problem-Oriented Medical Synopsis (POMS), based on work done over a twenty-year period. The conceptual schema of the system uses sub-models, and the record-keeping sub-model has been described elsewhere [1, 2]. Coding, indexing, and classification are critical components of clinical information systems, since a large part of the data in such systems consists of text, or narrative. Processing of "medical narrative" raises three fundamental issues. (i) How to **record** information: a standardized, preferred terminology is required, but such an instrument is not generally available for clinical problems; (ii) How to **locate** information selectively: this is done using a thesaurus of hierarchically organized nomenclature terms, such as SNOMed, the Systematized Nomenclature of Medicine; (iii) How to **summarize** information for analysis, using classification and aggregation. We have attempted to address this problem by combining the creation of domain-specific preferred terminologies with the modification of a classification system that is universal—ICD-9-CM, and a thesaurus of medical terms for indexing—SNOMed-III.

METHODS AND CONCLUSIONS

The preferred terminology we created is called the "problem glossary", and is strongly oriented towards general medicine and some of its sub-specialties, such as infectious disease, oncology, and pediatrics.

A biaxial classification and a "tangled hierarchy" were used to form a semantic network: there are about 600 codes in each axis. **Axis 1** is a body organ-system classification, a hybrid anatomic-physiologic organization. **Axis 2** is a health care process classification: "process" is defined as *that which is done in the course of the management of a reason for encounter, problem, or disease identified by a provider in a health care setting.*

The problem glossary has been reviewed by expert groups across multiple domains, and is thought to be comprehensive and easy to use. The content of the glossary was also cross-checked against the 5,000 most frequently occurring terms in MEDLINE and Excerpta Medica literature databases (MeSH and MALIMET thesauri)—about 3,000 of these are clinical terms, after excluding drug names and procedures. Each term is indexed with as many SNOMed-III terms as are needed. The problem glossary terms have been compared with the UMLS Metathesaurus terms. It seemed from this comparison that primary clinical terms (symptoms, signs, and test findings)may be underrepresented in the UMLS, at present.

A hybrid coding, indexing, and classification system has been created for medical record keeping and analysis of outcomes. The **content** is comprehensive for the domain of internal medicine and its sub-specialties, especially for diagnostic statements.

The **titles** are clear and unambiguous, and are based upon sound lexical rules. The **keys** are numeric (based upon a 4-byte integer), and are thus highly efficient for use in databases. The terms can be used for statistical **aggregation,** since each term is indexed using ICD-9-CM.

References

1. Stitt FW. Clinical database management for medical records. In: Cote RA, Protti DJ, Scherrer , JR, ed., *Role of informatics in health data coding and classification systems.* Amsterdam: North-Holland, 1985:295-331. IFIPS/IMIA conference proceedings.
2. Stitt FW. The Problem-Oriented Medical Synopsis: a patient-centered clinical information system. In: Safran C, ed., Annu Symp Comput Appl Med Care. New York: McGraw-Hill, Inc, Health Professions Division, November 1993:88-92.

0195-4210/94/$5.00